FIGURES 18A and 18B include flowcharts 310 and 410 that provide the series of logical steps employed to read reporter labeled beads in accord with the present invention. Referring to FIGURE 18A, in a start block 312, a first bead is read. In a block 313, spectral and/or spatial corrections are optionally applied to the pixelated signals from the detector to more accurately determine the spectral content in each pixel of the imagery. Next, the data from the bead imaging are segmented into a plurality of different reporter signature channels in a block 314. The number of signature channels in the present invention is chosen to be equal or greater than the number of different colors present in any reporter. In FIGURE 19, the reporter legend is based on reporters that comprise no more than 4 colors. Thus, FIGURE 17 includes four separate reporter signature channels, including "BLUE," "GREEN," "YELLOW," and "RED."

In a block 316, the deconvolution step discussed above is performed, but only if required, as would be the case for the first embodiment of the imaging and spectral decomposition. In a block 318 the binding signal channel is checked to determine if an analyte has bound to the bead indicating the presence of the analyte in the sample being tested. If no binding signal is present the logic determines if more beads are to be read in a decision block 324. If in decision block 324 it is determined that no more beads are to be read, a list of bead identities and binding strengths are generated in a block 326, as will be described in more detail below, with respect to FIGURE 18B. If in decision block 324 it is determined that more beads are to be read, the next bead is read in a block 328, and the logic loops back to block 313 as described above

If in decision block 318, it is determined that a binding signal is present, the presence and or strength of the binding signal is stored in a memory in a block 320. In a block 322, the bead is decoded as described in detail below in conjunction with FIGURE 18B. In the decoding process, when a valid bead signature is found, the signature or compound identity is stored along with the binding indicator. At this point the logic determines if more beads are to be read, in decision block 324, as described above.

The method for determining compound identity is illustrated by flowchart 410 in FIGURE 18B, which as noted above is executed in block 322 of FIGURE 18A, as noted in a block 420 in FIGURE 18A. After bead imagery has been segmented in each channel, individual reporters are evaluated to determine their identity. In a block 422 the imagery corresponding to each reporter is located in each signature channel. This may be accomplished by any means, including but no limited to, evaluating imagery in each signature channel to find the top left most reporter image and its corresponding imagery in other channels. Moving from left to right and top to bottom across the segmented region in each channel, every reporter is located in each channel. In a

block 424 a reporter is evaluated for its spectral content and the results are stored. In a decision block 428, the logic determines if more reporters are to be analyzed. If so, the logic loops back to block 424 to repeat the process for each reporter found in the image. Once it is determined in decision block 426 that no more reporters are to be analyzed, then in a block 428 the reporter signatures are compared against the reporter legend of FIGURE 19 to determine the compound identity.

In a decision block 430, the logic determines if the bead has a valid signature. If the signature is not valid, that occurrence is noted in a block 432. If the signature is valid, the signature is stored in a block 434. A signature is invalid if it is not supported by the reporter legend. The process illustrated in FIGURE 18B is repeated for each bead. It should be understood that when such a process is automated and executed by a computing device, the process of analyzing a single bead occurs very rapidly.

Flow Analysis of Libraries for DNA Sequencing, Polymorphism Analysis, and Expression Analysis

Given a bead library that comprises a complete set of DNA oligomers of length *N*, a subset of the library will hybridize to any piece of single stranded target DNA of approximately length *N* or greater. If the complementarity of the target DNA to a bead is less than *N* bases, hybridization strength (measured by the melting temperature, T_m) will drop. Six reporter colors employed in a binary encoding scheme are sufficient to encode a bead library of oligos up to length sixteen. As discussed above, the choice of oligo length affects the number of beads necessary for a complete library, the number of reporters necessary to encode the library, and the strength of hybridization. Shorter oligos require the smallest bead sets, the fewest reporters, and have the lowest tolerance for hybridization mismatch.

The shortest practical oligo length is approximately a 10-mer. A complete 10-mer library requires approximately one million beads, which can be analyzed in under an hour at a rate of a few hundred beads per second. The T_m of a perfect 10-mer complement is near room temperature, particularly for A/T rich sequences, so just a few base mismatches will likely cause the target DNA to dissociate from a bead at room temperature. Oligos of length 11 through 16 have melting temperatures significantly higher than room temperature and can therefore tolerate more mismatches than oligos of length 10, given the same reaction stringency. For the purposes of the following description, it will be assumed that a bead library of all possible 10-mers is used.

If a sample of target DNA consisting of multiple identical copies of length *M* is hybridized to a complete library, a subset of the beads numbering approximately *M*-10 will hybridize at random to the different strands of target DNA based on the sequences of their attached oligos. Each of the hybridizing oligos will have sequences that are

offset by one nucleotide, forming a contiguous overlapping string of sequences, as shown in FIGURE 20. Labeling the target DNA directly or indirectly with at least one luminescent probe will cause the bead to which the DNA is bound to emit light with the characteristic emission spectrum of the probe. If the probe spectrum is distinct from the spectra used for the reporters, the hybridization can be detected as shown in FIGURE 21.

After hybridization, the bead library is analyzed by flow imaging and the beads that emit a binding signal are decoded to reveal their associated oligonucleotide sequences. These sequences are then assembled into a "contig" of overlapping sequence by starting with a single oligo sequence and searching the remainder of the set of oligo sequences that bound to the target to find the closest match. The search can be highly constrained by the knowledge that all but a few of the bases near the end of the next oligo to be added to the contig must overlap the end sequence of the contig. The sequencing capacity of a bead library is constrained by the probability that the contig will branch and loop during the sequencing of a single stretch of DNA or that it will cross multiple fragments that are being sequenced simultaneously. The maximum length of a DNA sequence is approximately the square root of the size of the bead library. Hence, a complete 16-mer library of over four billion beads can sequence a stretch of DNA approximately 65,000 bases long.

As illustrated in FIGURE 20, the single-base offset of each oligo in the contig assures that there is *N*-fold sequence redundancy everywhere but at the ends of the contig. Further, the bead library can embody redundancy by including multiple copies of each unique bead (and corresponding oligo). Having multiple beads of each type enables the system to tolerate some fraction of unanalyzed beads due to system latency, defocus, out-of-view reporters, etc.

The sequencing protocol described herein can be implemented with a single target DNA sample or a pool of multiple targets, which will result in multiple contigs. The target DNA can be from any source, including genomic DNA, viral DNA, DNA produced by an amplification reaction such as a Polymerase Chain Reaction (PCR), or any other source known to those of ordinary skill in the art. In addition, DNA produced by an RNA template, RNA itself, or any other molecule known to hybridize to DNA, such as a peptide nucleic acid, can be used.

FIGURE 21 illustrates the application of the flow imaging sequencing method to determine the identity of SNPs, generally as indicated in the steps of FIGURE 22. A target DNA molecule is shown with a polymorphic site that can contain either a G or a T component. In sequencing the target, the ten beads that hybridize across the SNP site will either have a C or an A component at the SNP position if the source DNA is